

A Distributed Optimization Framework For Multi-channel Multi-user Small Cell Networks

Shuqin Li and Liyu Cai

Bell Labs China, Alcatel-Lucent Shanghai Bell Co., Ltd.

E-mail: {Shuqin.Li, Liyu.Cai}@alcatel-sbell.com.cn

Abstract—Small cell enchantment is emerging as the key technique for wireless network evolution. One challenging problem for small cell enhancement is how to achieve high data rate with as-low-as-possible control and computation overheads. As a solution, we propose a low-complexity distributed optimization framework in this paper. Our solution includes two parts. One is a novel implicit information exchange mechanism that enables channel-aware opportunistic scheduling and resource allocation among links. The other is the sub-gradient based algorithm with a polynomial-time complexity. What is more, for large scale systems, we design an improved distributed algorithm based on insights obtained from the problem structure. This algorithm achieves a close-to-optimal performance with a much lower complexity. Our numerical evaluations validate the analytical results and show the advantage of our algorithms.

I. INTRODUCTION

Wireless networks are undergoing a paradigm shift with the incoming new techniques to improve the capacity, including small cells (e.g. femtocells, picocells) and D2D (device-to-device) communications. It is likely that networks in a near future will become so unstructured that the traditional fully centralized control will be too complicated to be efficient. As a promising solution, a layering structure is recommended for heterogeneous networks according to the separation of data plane and control plane: the small cell layer, designed for high data rate transmission; and the macro cell layer, designed for large coverage and coordination among small cells. This paper focuses on the design problem of the small cell layer.

Compared to the traditional cells, small cells have many differences in terms of physical range, cost and functionality. First, the small cell has short transmit range. Thus simultaneous transmissions usually lead to strong interference. Second, to save the cost, small cells have rather limited processing capacities, in terms of both control power and computational power. Third, aimed for high data rates, the control overhead in small cell should be reduced as much as possible to save resource for data transmission. To tackle the new challenges brought along by these changes, we propose a distributed optimization framework in this paper, as a system solution for a synchronous multi-channel multi-user small cell network. In terms of low control and computational complexity, distributed systems always show great advantages over their centralized counterparts, not to mention they have scalability, flexibility and robustness for heterogeneous network environment, and also promise improvements in the utilization of the scarce spectrum resource. However, there are two main challenges

in the practical design of distributed network algorithms.

One is related to prohibitive overheads in channel information exchange. The traditional study in this area focuses on cellular networks (see [1] for a good tutorial), including one-to-one (UE to BS or BS to UE), or multiple-to-one (Multiple UEs to BS) feedbacks, which can be in either analog or digital way. To reduce the system overhead, partial information exchanges [1]–[3] are proposed to feedback information of only some (not all) channels, e.g., the best m channel mechanism. For a distributed system, where each link needs to transmit the channel information to all others in the network due to the lack of centralized control, the system overhead of information exchange can be very high even with the partial information exchange. To overcome this challenge, we design a novel *implicit* information exchange mechanism in our framework. It enables the channel-aware opportunistic scheduling and resource allocation by exploring the broadcast nature of wireless transmission: it encodes each link's channel information through the energy-levels of two simple signals, and enables links receiving these signals to inform the channel condition of others through simple calculations. This mechanism greatly reduces the system overhead.

The other challenge lies in rate maximization problem itself. Shown to be NP-hard [4] due to its non-convexity nature, this problem is extremely hard to solve even in a centralized way. To obtain the distributed solution, one popular technique is to reformulate this sum-rate-maximization problem as a non-cooperative game. Iterative Water Filling Algorithm (IWFA) [5] has been proposed to compute the Nash equilibrium of the game, and significant efforts [5]–[7] have been made to establish conditions to guarantee the IWFA convergence, the existence and the uniqueness of Nash equilibrium. However, in our study, we find that IWFA is not a favorable solution for small cell networks, mainly due to its unsatisfied performance under strong interference environment, not to mention its high complexity of iterative computation and insufficient resource allocation before the convergence. After all, a Nash equilibrium solution does not necessarily mean that it can be anywhere near the social optimality, e.g. the famous 'prison dilemma'. In contrast, we relax the rate maximization problem based on orthogonal resource allocation. We find that the concurrent transmissions in the small cell usually lead to strong interferences that greatly deteriorate the performance. Instead, the orthogonal resource allocation can achieve an optimal or near optimal throughput (shown by Hayashi and Luo in recent

paper [8]). In our framework, we design distributed scheduling and power allocation algorithms with low complexity, which achieves high data rates and supports QoS and fairness among multi-users.

II. PROBLEM FORMULATION

We consider a general multi-channel multi-user network with $\mathcal{I} \triangleq \{1, 2, \dots, I\}$ links¹ sharing a common spectrum which is divided into $\mathcal{K} \triangleq \{1, 2, \dots, K\}$ frequency tones. With a little abuse of notation, we use the same index of link to denote its corresponding transmitter and receiver. Thus G_{ij}^k denotes the channel coefficient² between link i 's transmitter and link j 's receiver in tone k , and N_0 is the noise variance. The transmit power of link i in tone k is denoted as p_i^k , and r_i denotes link i 's transmit rate, which is the sum of link i 's rates in all tones. Therefore, the weighted sum Rate Maximization (RM) problem for the system can be formulated as follows:

$$\begin{aligned} \text{RM : Maximize} \quad & \sum_{i \in \mathcal{I}} \theta_i r_i \\ & p_i^k \geq 0 \\ \text{Subject to} \quad & \sum_{k \in \mathcal{K}} p_i^k \leq P_{i0}, \forall i \in \mathcal{I} \end{aligned} \quad (1)$$

where θ_i is the weight of each link i , a parameter predetermined by the system to enable some design flexibility (e.g., priorities, fairness, and QoS requirements etc.). The constraint in (1) denotes that each link i has a limited transmit power of each link, i.e., no larger than its maximum power P_{i0} .

We apply orthogonal resource allocation (i.e., tones are orthogonally shared by links) in our solution. Thus the transmit rate of link i can be calculated according to Shannon formulation $r_i \triangleq \sum_{k \in \mathcal{K}} \log(1 + g_i^k p_i^k)$, where $g_i^k \triangleq \frac{G_{ii}^k}{N_0}$ is the normalized channel gain for link i in tone k .

There are two technical challenges to solve the RM problem in a distributed manner. First, the channel gain g_i^k is private information of each link. Thus a simple and effective information exchange mechanism is needed. Second, the RM problem has been shown to be NP-hard [4]. Thus an efficient algorithm is needed to compute an approximated solution with satisfactory performances.

III. THE OPTIMIZATION FRAMEWORK

We propose a distributed optimization framework to tackle the aforementioned challenges. This framework includes two key functional modules: one *signaling* module that enables channel information exchange between links, and the other *scheduling and resource allocation* module that solves the RM problem. Operations of these functional modules in the framework will run in a time-slot base. The timeline of the framework is illustrated in Fig. 1. In the section III-A and III-B, we will discuss the design of each module in more details, respectively.

¹A link is composed by a pair of transmitter and its receiver, which can be a D2D pair with two User Equipments (UE), or an uplink pair between a UE and the evolved Node B (eNB).

²If we define the received signal as $y_i^k = H_{ii}^k x_{ii}^k + z_i$, then $G_{ii}^k = (H_{ii}^k)^2$, and N_0 is the noise z_i 's variance.

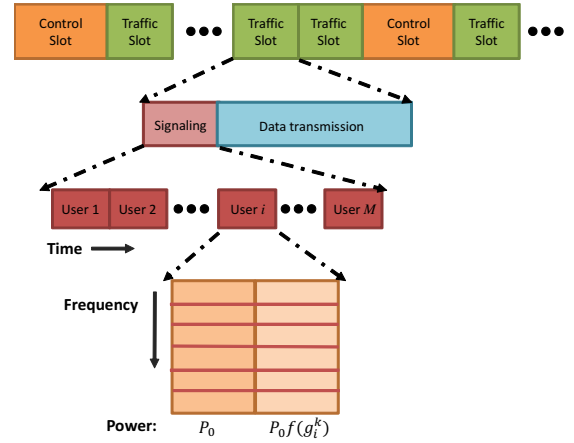


Fig. 1. System Timeline: each traffic slot is composed by a signaling slot and a transmission slot, and each signaling slot is further divided into M sub-slots. Each link will be allocated such a sub-slot to do two signal transmissions.

A. Inexplicit Signaling

In the signaling module, we design an inexplicit information exchange mechanism that enables links to coordinate in scheduling and resource allocation to achieve a good system performance in a distributed manner. Specifically, this mechanism encodes the channel condition of each link through the energy-levels of two simple signals, and enables each link to inform the channel condition of others through simple calculations. It works in the signaling slot of each traffic slot as shown in Fig. 1. Each signaling slot is further divided into M sub-slots, where $M \geq I$ guarantees that each link in the network can be allocated one sub-slot. The allocation can follow links' predetermined indices (e.g., according to their priorities). In the signaling slot, links sequentially inform their channel gain in their own sub-slot as following two steps:

- Step 1: each link i broadcasts a signal with power P_0 in each tone k .
- Step 2: each link i broadcasts a signal with power $P_0 f(g_i^k)$ in each tone k .

$f(g_i^k)$ is public information of links, a function predetermined by the system, which can be any monotonically increasing function of the channel gain g_i^k . For example, a nice choice of $f(g_i^k)$ can be the cumulative distribution function (CDF) of the channel gain g_i^k . When another link, say a link j , hears these two signals, denoted as S_1 and S_2 , it can quickly obtain the channel gain g_i^k by the following simple calculation:

$$g_i^k = f^{-1} \left(\frac{S_2}{S_1} \right) = f^{-1} \left(\frac{g_{ij}^k P_0 f(g_i^k)}{g_{ij}^k P_0} \right) \quad (2)$$

It is more convenient to adopt discrete function of $f(g_i^k)$ for implementation, since each g_i^k is also a sampled and quantified discrete value in practical systems. Thus the values of the function $f(g_i^k)$ are stored as a table, and the operation in (2) is simply a table searching.

We can take the following simple example to see this process in a more clear way. Suppose that the channel gains are quantified as three values: HIGH, MIDDLE, LOW, and

the system can simply predetermine

$$f(g_i^k) = \begin{cases} 1 & \text{if } g_i^k = \text{HIGH}, \\ \frac{2}{3} & \text{if } g_i^k = \text{MIDDLE}, \\ \frac{1}{3} & \text{if } g_i^k = \text{LOW}. \end{cases} \quad (3)$$

Next, when the link j obtains the power quotient of two received signals, say, $2/3$, then it immediately infers that the channel gain of g_i^k is the MIDDLE value.

In practical implementation, we can choose much denser samplings for quantifying channel conditions g_i^k (e.g., far more than three channel conditions in this toy example). Thus the more accurate channel conditions, the more benefits for computing a better scheduling and resource allocation in the transmission module.

B. Scheduling and Resource Allocation

After the signaling slot, each link knows the channel information of all other links. The transmission module of each link will individually calculate the scheduling and resource allocation solution to the RM problem in the next data transmission slot as shown in Fig. 1. We first propose a sub-gradient algorithm which converges to the optimal value guaranteed by strict mathematical proofs. Based on the insights obtained by this algorithm, we then proposed a more efficient sub-optimal solution with much lower complexity, and yet a close-to-optimal performance.

1) *Optimal Subgradient Algorithm*: The RM problem is shown to be NP-hard. To see it clearly, we introduce binary variables T_i^k : tone k is assigned to link i , if $T_i^k = 1$; otherwise $T_i^k = 0$. The transmission rate is $r_i \triangleq \sum_{k \in \mathcal{K}} T_i^k \log \left(1 + \frac{g_i^k p_i^k}{T_i^k} \right)$. Since the introduction of T_i^k does not change the value of transmission rate, the resulting problem is equivalent to the RM problem. We can see that the NP-hardness of the RM problem roots in the combinatorial choice of channel assignment, i.e., the binary T_i^k . To resolve this difficulty, we relax the binary variable T_i^k into the range from 0 to 1, i.e., $T_i^k \in [0, 1]$. The corresponding physical meaning is that each channel can be Time Sharing (TS) among several links:

$$\begin{aligned} \text{TS: } & \text{Maximize} \quad \sum_{i \in \mathcal{I}} \theta_i r_i \\ & \text{Subject to} \quad \sum_{k \in \mathcal{K}} p_i^k \leq P_{i0}, \forall i \in \mathcal{I} \quad (4) \\ & \quad \quad \quad \sum_{i \in \mathcal{I}} T_i^k \leq 1, \forall k \in \mathcal{K} \quad (5) \end{aligned}$$

We can show that the TS problem is a convex optimization problem, and that there is no duality gap. Thus we can solve the equivalent dual problem by subgradient algorithms [10]. Due to the page limit, technical details are provided in Appendix A and B of our technical report.

2) *Low Complexity Suboptimal Algorithm*: There are several major difficulties to apply the subgradient algorithm in the practical system. First, despite its polynomial complexity, the subgradient search still converges too slowly to be useful for the scheduling on a fast scale. Second, the time-sharing

result will cause overhead of synchronizing links to share data transmission slot together. However, by using the structure information revealed by the subgradient algorithm, we propose the following more efficient Sub-Optimal Algorithm (SOA).

In the subgradient solution, given the optimal dual variable λ^* , the channel assignment is determined by sorting the links on each tone according to a complex metric ξ_i^k , and then given the optimal channel assignment, the power allocation is given by a water filling solution (refer to Appendix B of our technical report). In the SOA, we adopt the same two phases, but modify them to reduce the complexity.

Channel Assignment Phase: In this phase, each channel will be assigned to at most one link. Instead of ξ_i^k , we introduce a new metric (6) based on the equal power allocation over all tones assigned to a link. The pseudo code is given in Algorithm 1. In this algorithm, *No-Assigned-Channel* is a

Algorithm 1 Channel Assignment for SOA

- 1: Initialization: *No-Assigned-Channel* = 0,
 Assigned-Channel-Set _{i} = \emptyset
 - 2: **while** *No-Assigned-Channel* < K **do**
 - 3: Update the index *Channel-To-Assign* _{i} for each link i
 - 4: Update the metric *Marginal-Rate* _{i} for each link i
 - 5: Find $i^* \leftarrow \arg \max \{ \textit{Marginal-Rate}_i > 0, i \in \mathcal{I} \}$
 - 6: Assign the channel *Channel-To-Assign* _{i^*} to link i^* by updating:
 No-Assigned-Channel \leftarrow *No-Assigned-Channel* + 1
 Assigned-Channel-Set _{i^*} \leftarrow *Assigned-Channel-Set* _{i^*} \cup *Channel-To-Assign* _{i^*}
 - 7: **end while**
-

counter denoting how many tones have been assigned, and *Assigned-Channel-Set* _{i} denotes the tones assigned to link i . The index *Channel-To-Assign* _{i} denotes the tone to be assigned to the link i in each iteration. It is determined as follows: we sort the tones based on the channel gain for each individual link. For each link i , *Channel-To-Assign* _{i} is the tone index with the largest channel gain in the unassigned channels. The metric *Marginal-Rate* _{i} is defined as the marginal rate if the channel *Channel-To-Assign* _{i} is allocated to the link, i.e.,

$$\begin{aligned} \textit{Marginal-Rate}_i = & \theta_i \sum_{k \in \textit{ACS}_i \cup \textit{CTA}_i} \log \left(1 + \frac{P_{i0} g_i^k}{|\textit{ACS}_i| + 1} \right) \\ & - \theta_i \sum_{k \in \textit{ACS}_i} \log \left(1 + \frac{P_{i0} g_i^k}{|\textit{ACS}_i|} \right) \end{aligned} \quad (6)$$

where \textit{ACS}_i is the abbreviation of *Assigned-Channel-Set* _{i} , and \textit{CTA}_i is the abbreviation of *Channel-To-Assign* _{i} .

Power Allocation Phase: After the channel assignment, we can adopt equal power allocation, which removes all computations of the water filling solution. However, in the simulation, we find that it can maintain almost the same performance of water filling solution.

It is not hard to see that the total complexity of SOA is $\mathcal{O}(IK \log(K))$.

IV. PERFORMANCE

The performance of the proposed framework is examined in this section. Several different indoor small cell scenarios (urban or suburban, inside or outside door, and also different radius) are considered and tested. Due to the page limit, we only show results for one typical scenario. For other scenarios and discussions on implementation issues, readers can refer to Appendix C and D of our technical report.

Here we consider a urban indoor small cell scenario, where different number (2–10) of links are uniformly distributed (*i.e.*, both transmitters and receivers are randomly dropped) in a small cell with radius of 25m, sharing 10 different tones, with 180kHz for each tone³. The pathloss model is $PL(dB) = 38.46 + 20 \log_{10} R + 0.7d_{2D,indoor} + 18.3n^{(n+2)/(n+1)-0.46} + q L_{iw}$ with $\ln H(LOS)$ fading⁴. The maximum power of each link is $P_{i0} = 20$ dBm, $\forall i \in \mathcal{I}$. The noise power density is $N_0 = -174$ dBm/Hz.

Two popular algorithms are compared with our algorithm (SOA). One is the distributed algorithm IWFA. The weights are set as $\theta_i = 1$, $i \in \mathcal{I}$, so that IWFA can work. The other is MAPLE [9], which is a centralized algorithm, proved to be throughput-optimal for single channel scenario. In our simulation, we extend it for the multi-channel scenario by averagely splitting the link power budget, independently running MAPLE in each tone, and taking the sum rate of all tones as its throughput. Fig. 2 shows the throughputs of these three algorithms as the number of links increases, in which each point in each curve is with 100 trials. Tab. I records the average execution time of one trial for three algorithms⁵.

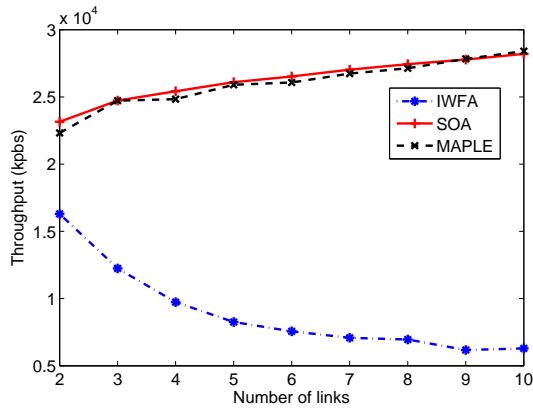


Fig. 2. Performance comparisons of three algorithms

As shown in Fig. 2, the proposed SOA has a very similar (sometimes even better) throughput as MAPLE. Their perfor-

³A typical system would be the 20MHz LTE system: 10 resource blocks, 180kHz for each, with 2MHz for guard band.

⁴Parameter $d_{2D,indoor}$ is the distance inside the room. n is the number of penetrated floors. q is the number of walls, L_{iw} is the penetration loss, 5dB. Readers can refer to 3GPP's standardization 36.814 for more explanations. The result in this letter use the settings $d_{2D,indoor} = 25m$, $n = 0$, $q = 1$. We test other values of these parameters, the value of throughputs will be different, but the trends of curves are similar.

⁵All trails are run in exactly the same computational environment. In Tab. I, we use the following abbreviations for MAPLE's average execution time: $1.80e3 = 1.80 \times 10^3$, $7.71e3 = 7.71 \times 10^3$ and $3.24e4 = 3.24 \times 10^4$.

No. of link	2	3	4	5	6	7	8	9	10
SOA (μs)	32.6	32.3	31.1	29.7	29.4	29.9	29.5	30.1	29.6
IWFA (ms)	0.36	0.59	0.90	1.27	1.72	2.21	2.79	3.61	4.26
MAPLE (s)	0.06	0.22	0.75	2.56	11.5	55.8	1.80e3	7.71e3	3.24e4

TABLE I
EXECUTION TIME OF THREE ALGORITHM

mances are increasing with the number of links increasing, which is mainly due to the multi-use diversities in the frequency domain. We observe that the performance of IWFA deteriorates quickly as the number of links increases. Compared with it, SOA has a large performance gain, from 142% for 2 links to 449% for 10 links. As we know, spatial diversity gain and interference are two sides of one coin for concurrent transmissions. Designed to explore the spatial diversity, IWFA tries to enable simultaneous transmissions as many as possible, which is in general not suitable for such a strong interference environment as small cells as shown in our simulation, for the performance loss caused by interference greatly overweighs the gain possibly brought by spatial diversity.

In terms of system complexity of algorithms, there are two parts, control complexity and computational complexity. For the first part, it highly relates to the concrete system implementation. However, several things are for sure: First, the inexplicit signaling mechanism proposed in Sec. III-A has great advantage for reducing feedback overheads, and can be customized to work with other resource allocation algorithms. Second, enabling by the inexplicit signaling, SOA can independently work on each link, and obtain the resource allocation result by one calculation as the centralized algorithms, *e.g.*, MAPLE. However, IWFA usually takes dozens of iterations (*i.e.*, one calculation and resource allocation per link per iteration) before its convergence. For the computational complexity, the comparisons will be more direct, just by checking the execution time in Tab. I, which is proportional to the algorithm's computational complexity. It is clear that SOA has a very low complexity.

V. CONCLUSION

In this paper, we propose a distributed optimization framework for the multi-channel multi-link small cell network, which enables inexplicit information exchange and channel-aware opportunistic scheduling and resource allocation. It achieves a close-to-optimal performance with very low complexity, and thus shows great potentials for small cell applications in future heterogeneous networks.

REFERENCES

- [1] D. Love, R. Heath, V. Lau, D. Gesbert, B. Rao, D. Bhaskar and M. Andrews, "An overview of limited feedback in wireless communication systems", *IEEE Journal on Selected Areas in Communications*, 2008.
- [2] Y. Huang and B. Rao, "Sum rate analysis of one-pico-inside OFDMA network with opportunistic scheduling and selective feed-back", *IEEE Signal Process. Lett.*, 2012.
- [3] Y. Huang and B. Rao, "An analytical framework for heterogeneous partial feedback design in heterogeneous multicell OFDMA networks", *IEEE Transactions on Signal Processing*, 2013.
- [4] Z. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE Journal of Selected Topics in Signal Processing*, 2008.
- [5] W. Yu, G. Ginis, and J. Cioffi, "Distributed multiuser power control for digital subscriber lines," *IEEE JSAC* 2002.

- [6] G. Scutari, D. Palomar, and S. Barbarossa, "Optimal linear precoding strategies for wideband noncooperative systems based on game theory-part i: Nash equilibria," *IEEE Transactions on Signal Processing*, 2008.
- [7] Z. Luo and J. Pang, "Analysis of iterative waterfilling algorithm for multiuser power control in digital subscriber lines," *EURASIP Journal on Advances in Signal Processing*, 2006.
- [8] S. Hayashi and Z. Luo, "Spectrum management for interference-limited multiuser communication systems," *IEEE Transactions on Information Theory*, 2009.
- [9] L. Qian, Y. Zhang and J. Huang, "MAPEL: Achieving global optimality for a non-convex wireless power control problem", *IEEE Transactions on Wireless Communications*, vol. 8, no. 3, pp.1553–1563, 2009
- [10] D. P. Bertsekas, *Nonlinear Programming*. Athena Scientific, 1999.

APPENDIX

A. Derivation of Sub-gradient Algorithm

The dual problem of the TS problem is shown as follows:

$$\text{Minimize } D(\boldsymbol{\mu}, \boldsymbol{\lambda}), \quad (\mu, \lambda) \geq 0 \quad (7)$$

where λ_i and μ^k , $i \in \mathcal{I}$, $k \in \mathcal{K}$ denote the Lagrange multipliers associated with constraints (4) and (5), and the bold font denotes the vector form of the corresponding variable. The dual function is defined as

$$D(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \max_{\mathbf{p} \geq 0, \mathbf{T} \in [0,1]} L(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{p}, \mathbf{T}). \quad (8)$$

and the Lagrangian is defined as

$$L(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{p}, \mathbf{T}) = \sum_{i \in \mathcal{I}} \theta_i r_i - \sum_{k \in \mathcal{K}} \mu^k \left(\sum_{i \in \mathcal{I}} T_i^k - 1 \right) - \sum_{i \in \mathcal{I}} \lambda_i \left(\sum_{k \in \mathcal{K}} p_i^k - P_{i0} \right).$$

We first calculate the dual function as follows. For given $\boldsymbol{\mu}, \boldsymbol{\lambda}$, we obtain the optimal \mathbf{p}^* by maximizing the Lagrangian $L(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{p}, \mathbf{T})$ through the first-order derivative condition:

$$p_i^{k*} = \frac{T_i^k}{h_i^k} \left[\frac{\theta_i h_i^k}{\lambda_i} - 1 \right]^+, \quad \forall i \in \mathcal{I}, k \in \mathcal{K}, \quad (9)$$

where we use the notation $(\cdot)^+ \triangleq \max(\cdot, 0)$. With the optimal \mathbf{p}^* , the resulting Lagrangian is

$$L(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{p}^*, \mathbf{T}) = \sum_{i \in \mathcal{I}} \sum_{k \in \mathcal{K}} (\xi_i^k(\lambda_i) - \mu^k) T_i^k + \sum_{k \in \mathcal{K}} \mu^k + \sum_{i \in \mathcal{I}} \lambda_i P_{i0},$$

where

$$\xi_i^k(\lambda_i) \triangleq \begin{cases} \theta_i \left[\log \left(\frac{\theta_i h_i^k}{\lambda_i} \right) - \frac{\lambda_i}{\theta_i h_i^k} + 1 \right] & \text{If } \theta_i h_i^k > \lambda_i, \\ 0 & \text{Otherwise.} \end{cases} \quad (10)$$

It is not hard to see that $L(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{p}^*, \mathbf{T})$ is linear in \mathbf{T} . Thus we quickly obtain the optimal value \mathbf{T}^* by maximizing $L(\boldsymbol{\mu}, \boldsymbol{\lambda}; \mathbf{p}^*, \mathbf{T})$:

$$T_i^{k*} = \begin{cases} 1, & \text{If } \xi_i^k(\lambda_i) > \mu^k, \\ \text{any value} \in [0, 1], & \text{If } \xi_i^k(\lambda_i) = \mu^k, \\ 0, & \text{If } \xi_i^k(\lambda_i) < \mu^k. \end{cases} \quad (11)$$

With the optimal \mathbf{p}^* and \mathbf{T}^* , we obtain the dual function as

$$D(\boldsymbol{\mu}, \boldsymbol{\lambda}) = \sum_{k \in \mathcal{K}} \sum_{i \in \mathcal{I}} (\xi_i^k(\lambda_i) - \mu^k)^+ + \sum_{k \in \mathcal{K}} \mu^k + \sum_{i \in \mathcal{I}} \lambda_i P_{i0} \quad (12)$$

Next we solve the dual problem (7) by minimizing the dual function (12) over $\boldsymbol{\mu}$ for given $\boldsymbol{\lambda}$.

We can show that the optimal value of $\boldsymbol{\mu}$ is given by

$$\mu^{k*} = \max_{i \in \mathcal{I}} \{\xi_i^k\} \triangleq \xi_{(1)}^k, k \in \mathcal{K}. \quad (13)$$

where $\xi_{(1)}^k$ denotes the largest elements of the set $\{\xi_i^k\}_{i \in \mathcal{I}}$. Substituting all above results, we have the equivalent dual problem, shown as follows:

$$\text{Minimize } D(\boldsymbol{\lambda}) \quad \boldsymbol{\lambda} \geq 0 \quad (14)$$

where $D(\boldsymbol{\lambda}) \triangleq \sum_{k \in \mathcal{K}} \xi_{(1)}^k(\lambda_i) + \sum_{i \in \mathcal{I}} \lambda_i P_{i0}$.

Problem (14) is a single variable convex optimization problem. To solve it, we adopt a subgradient-based search and update $\boldsymbol{\lambda}$ as follows:

$$\boldsymbol{\lambda}(t+1) = \left[\boldsymbol{\lambda}(t) - \alpha(t) \left(\mathbf{P}_0 - \sum_{k \in \mathcal{K}} \mathbf{p}^{k*}(t) \right) \right]^+ \quad (15)$$

where \mathbf{p}^{k*} are given by (9), $\alpha(t)$ is the step size. By choosing a diminishing step size that is *square summable but not summable*, the subgradient algorithm can be proved to converge to the optimal solution within polynomial time [10]. We give the proof in the next section.

B. Convergence of Sub-gradient Algorithm

Proof: Recall that a subgradient of f at x is any vector v that satisfies the inequality $f(y) \geq f(x) + v^\top(y - x)$ for all y . By this definition, it is not hard to show that

$$\mathbf{g} \triangleq \mathbf{P}_0 - \sum_{k \in \mathcal{K}} \mathbf{p}^{k*}$$

is the subgradient of function $D(\boldsymbol{\lambda})$ at $\boldsymbol{\lambda}$.

Then we have

$$\begin{aligned} \|\boldsymbol{\lambda}(t+1) - \boldsymbol{\lambda}^*\|_2^2 &= \|\boldsymbol{\lambda}(t) - \alpha(t)\mathbf{g}^\top(t) - \boldsymbol{\lambda}^*\|_2^2 \\ &= \|\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*\|_2^2 - 2\alpha(t)\mathbf{g}^\top(t)(\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*) + \alpha^2(t)\|\mathbf{g}(t)\|_2^2 \\ &\leq \|\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*\|_2^2 - 2\alpha(t)(D(\boldsymbol{\lambda}(t)) - D(\boldsymbol{\lambda}^*)) + \alpha^2(t)\|\mathbf{g}(t)\|_2^2. \end{aligned}$$

Applying the inequality above recursively, it follows

$$\begin{aligned} \|\boldsymbol{\lambda}(t+1) - \boldsymbol{\lambda}^*\|_2^2 &\leq \|\boldsymbol{\lambda}(1) - \boldsymbol{\lambda}^*\|_2^2 \\ &\quad - 2 \sum_{i=1}^t \alpha(i)(D(\boldsymbol{\lambda}(i)) - D(\boldsymbol{\lambda}^*)) + \sum_{i=1}^t \alpha^2(i)\|\mathbf{g}(i)\|_2^2. \end{aligned}$$

By the fact $\|\boldsymbol{\lambda}(t) - \boldsymbol{\lambda}^*\|_2^2 \geq 0$ and $\|\boldsymbol{\lambda}(1) - \boldsymbol{\lambda}^*\|_2^2 \leq R$, where R is a predetermined bound of $\boldsymbol{\lambda}$, we have

$$2 \sum_{i=1}^t \alpha(i)(D(\boldsymbol{\lambda}(i)) - D(\boldsymbol{\lambda}^*)) \leq R^2 + \sum_{i=1}^t \alpha^2(i)\|\mathbf{g}(i)\|_2^2$$

Combining this with

$$\begin{aligned} &\sum_{i=1}^t \alpha(i)(D(\boldsymbol{\lambda}(i)) - D(\boldsymbol{\lambda}^*)) \\ &\geq \sum_{i=1}^t \alpha(i) \min_{i=1, \dots, t} D(\boldsymbol{\lambda}(i)) - D(\boldsymbol{\lambda}^*) \\ &= \sum_{i=1}^t \alpha(i) D(\boldsymbol{\lambda}_{best}) - D(\boldsymbol{\lambda}^*) \end{aligned}$$

and by the assumption $\|\mathbf{g}(i)\|_2^2 \leq G^2$, we have

$$0 \leq D(\boldsymbol{\lambda}_{best}) - D(\boldsymbol{\lambda}^*) \leq \frac{R^2 + G^2 \sum_{i=1}^t \alpha^2(i)}{\sum_{i=1}^t \alpha(i)} \quad (16)$$

Therefore, when α is *square summable but not summable*, i.e., $\|\alpha\|_2^2 = \sum_{i=1}^t \alpha^2(i) < \infty$, and $\sum_{i=1}^t \alpha(i) = \infty$, we have the convergence results. ■

scenario		Pathloss
1) Urban	UE insider the door	$38.46 + 20 \log_{10} R + 0.7 d_{2D, \text{indoor}} + 18.3 n^{\frac{n+2}{n+1}-0.46} + q L_{iw}$
2) Urban	UE outsider the door	$\max(38.46 + 20 \log_{10} R, 15.3 + 37.6 \log_{10} R) + 0.7 d_{2D, \text{indoor}} + 18.3 n^{\frac{n+2}{n+1}-0.46} + q L_{iw} + L_{ow}$
3) Suburban	UE insider the door	$38.46 + 20 \log_{10} R + 0.7 d_{2D, \text{indoor}} + 18.3 n^{\frac{n+2}{n+1}-0.46}$
4) Suburban	UE outsider the door	$\max(38.46 + 20 \log_{10} R, 15.3 + 37.6 \log_{10} R) + 0.7 d_{2D, \text{indoor}} + 18.3 n^{\frac{n+2}{n+1}-0.46} + L_{ow}$

TABLE II
PATHLOSS MODEL FOR DIFFERENT INDOOR SMALL CELL SCENARIO

C. More simulation results for different scenarios

The simulation setting is the same as the one in Sec. IV. We have four difference pathloss model⁶ shown in Tab. II. The first scenario in Tab. II is the one in Sec. IV.

For each scenario, we test the cell radius from 10m to 40m. We observe that the trends of the these curves are rather similar as Fig. 2, therefore three other results for different scenarios and different cell radius are picked here, as shown in Fig. 3, Fig. 4 and Fig. 5. Readers can refer to Sec. IV for performance analysis.

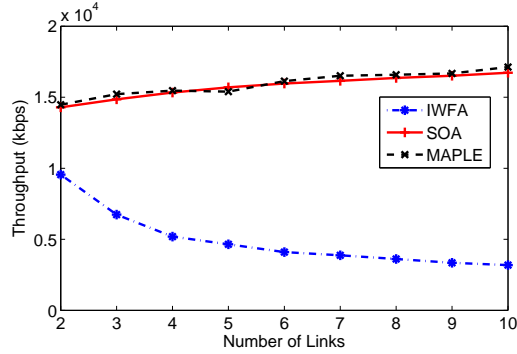


Fig. 3. Scenario 2) with 10m cell radius.

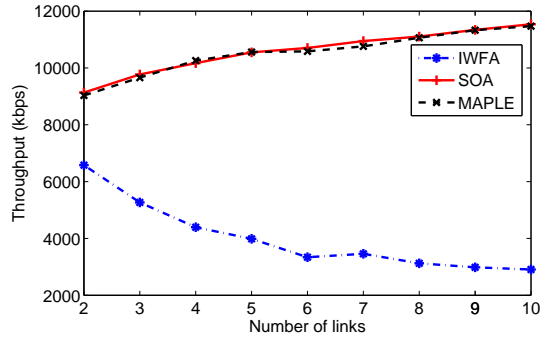


Fig. 4. Scenario 3) with 30m cell radius.

D. Discussions of Implementation-related Issues

In this section, we give some discussions of implementation-related issues for the proposed framework in practical system.

1) *About Signaling Errors:* What happens if some links fail to receive the signaling from other links in the signaling slot? Does the scheduling of the resource allocation algorithms in

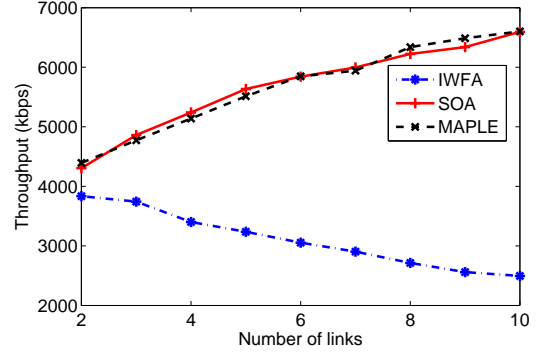


Fig. 5. Scenario 4) with 40m cell radius.

the transmission module still work? We discuss this problem in two cases:

Case 1: There are very weak interferences or almost no interference between some links, so they cannot receive signals from each other. For example, there is a long enough distance between them. Since they are just transparent to each other, there is no problem that each of them runs the proposed framework. It is actually a good case, since it can explore the spatial diversity.

Case 2: There are substantial or strong interferences among links, but somehow some of links, say a link *A*, loses some signals from others in some channel, say a link *B*'s signal in a channel *c*. Then link *A* would think that channel *c* is a bad channel for link *B* (since the channel gain is near 0). For this scenario, the scheduling and resource allocation algorithms still work, but there may be collisions in the channel allocation, and the data rate of each link decreases due to the interference. For example, both link *A* and link *B* choose to transmit in channel *c*. How to solve this collision issue? First, the collision is easy to detect after it happens: the data rate in the collision channel is not as high as the algorithm calculated. Once a link knows there is a collision happened in one tone, there are many ways to solve the problem. A simple one may be that the link gives up the tone with some probability and recalculates scheduling and resource allocation for the remaining channels in the next slot.

In summary, the proposed framework is robust to errors in signaling, but there will be performance loss in case 2. How much is the loss, depends on how often the error happens and how we design links to avoid collisions when a collision happened (e.g., the probabilities to give up tones in the aforementioned mechanism). But since the proposed framework is designed for a small cell network, where links are located in a small range of (usually indoor) area, both case

⁶Here L_{ow} is the penetration loss of an outdoor wall, 20dB.

1 and case 2 rarely happen.

2) *About Control Overheads:* Major control overhead comes from signaling module. How often we should do signaling depends on how fast the channels change: the slower the channel variations, the less often the signaling, and thus the less overhead. The proposed framework is designed for the small cell network scenario, which will be deployed in hot spots with low mobility where the channel quality only varies slowly with time.

Secondly, the proposed framework is able to work for general multi-channel system, since we do not have any limits on the physical layer technology. And there will also have many different implementation ways when we customize the signaling mechanism into a particular multi-channel system, since we leave many design freedoms for the signaling mechanism. In a nutshell, the signaling scheme just requires two transmissions with particular power levels for each link in each channel. To decrease the overhead, the length of the sub-slot can be very short, as long as that the power strength can be detected by other users in the sub-slot.

Further, we can also add channel prediction algorithms to further decrease the signaling frequency. It decreases the system overhead but may increase the estimation error. It is important to choose a proper tradeoff of frequencies of running signaling and channel prediction algorithm according to the network environment.